近年来知识融合研究进展与趋势

■ 朱祥 张云秋
吉林大学公共卫生学院医学信息学系 长春 130021

摘要：[目的/意义]对近年来知识融合相关研究进行梳理与评价，以期为今后相关研究提供参考。[方法/过程]首先解析知识融合的概念，然后对知识融合的框架、过程和方法进行梳理，继而总结知识融合的研究趋势，最后进行研究展望。[结果/结论]知识融合研究在大数据环境下呈现出新的研究特点，但还不能满足大数据环境的要求，未来应从构建分层多维立体的大数据知识融合框架、提高知识融合的效率、构建实时动态融合机制、开展大数据实证应用研究4个方面开展知识融合研究。

关键词：知识融合 大数据 述评
分类号：G250
DOI:10.13266/j.issn.0252-3116.2019.16.015

随着大数据时代的到来，数据来源更加广泛，数据结构更加复杂多样。数据的多源异构是大数据时代进行"全数据"分析面临的最主要的问题，而知识融合是解决数据多源异构问题的有效途径，可以实现语义层次上对数据的深度加工，形成新的知识，提供新服务。知识融合在大数据时代愈发重要，也成为近年来情报学领域新的关注点。事实上，知识融合研究刚刚起步，成熟的研究尚少，因此有必要对目前知识融合相关研究进行及时梳理与评价，提出未来的研究方向，以期对今后的知识融合研究提供参考。本文以"knowledge fusion"和"知识融合"为检索词分别在Web of Science核心合集和CNKI进行主题检索，时间范围为2012年1月至2018年9月，以获取大数据元年以来发表的相关文献。经过逐一阅读筛选，共得到55篇密切相关文献，对这些文献进行进一步追溯，又获得13篇相关文献，共计68篇，这些文献构成了本研究的基础。本文首先对知识融合的概念进行了解析，然后按照"框架-过程-方法"的逻辑顺序，分别对知识融合的框架、过程和方法3个关键方面进行了梳理，继而总结了知识融合的研究趋势，最后面向大数据环境进行了研究展望。

1 知识融合的概念解析

知识融合(knowledge fusion, KF)是由数据融合、信息融合发展而来的。数据融合(data fusion, DF)最早起源于20世纪70年代的军事领域，主要是综合利用来自多个传感器的数据，形成对目标的综合判断[1]。90年代，信息融合(information fusion, IF)的概念开始出现，与数据融合相比，信息融合的处理对象扩展到了多源信息。它是对多源信息进行综合处理的过程，以期得到比单一信息源更精确、更可靠的推理决策[2]。90年代末，知识融合开始得到关注。在图书情报学、管理学、计算机科学等学科中，知识融合位于知识科学视角下，知识融合的对象已经不局限于传感器获取的信息，而是已有的知识库，或者从数据库、网络、业务系统中抽取而来的知识，甚至扩展到各种方法、专家经验等。有学者认为三者在研究对象、处理单元、融合过程和融合结果上都有很大程度的交叉[3]，也有学者认为知识融合包括了数据融合和信息融合[4]。由于数据、信息，也是大数据环境下知识融合的处理对象和产物，且数据、信息和知识三者具有转化关系，因此本文认为知识融合包括了数据融合与信息融合。

在知识融合的发展过程中，存在两种观点的定义：第一种定义强调知识融合是将抽取到的多源知识转换为统一知识模式的过程；而第二种定义更强调融合结果，表明知识融合不是对多源异构数据源的简单整合或集成，其最显著的特征是有新知识的产生，这种观点在近年来的大数据环境下被普遍接受。综合研究者

作者简介：朱祥(ORCID:0000-0002-5608-5272)，博士研究生；张云秋(ORCID:0000-0002-9790-9581)，信息检索教研室主任，教授，博士生导师，通讯作者，E-mail:yunqiu@jlu.edu.cn。
收稿日期:2018-10-25 修回日期:2019-03-14 本文起止页码:143-150 本文责任编辑:易飞

们[5-8]提出的定义，本文认为大数据环境下的知识融合是面向知识服务和决策支持，以海量的多源异构数据、信息、知识为基础，利用融合算法和规则，在语义层次上组合、推理、创造出新知识的过程。

2　知识融合框架

知识融合框架是实现知识融合的基础。知识融合框架研究一直以来都是知识融合研究的重点，框架一般集成了最新的理论方法，而框架研究指导了知识融合方法和应用研究。本部分从宏观和微观两个层面对知识融合框架进行梳理：宏观上可以分为以方法为导向和以服务为导向的知识融合框架；微观上又可以分为横向和纵向两个视角下的知识融合框架，其中纵向即分层融合，横向即多维融合。最后梳理了专门面向大数据环境提出的知识融合框架。

2.1　宏观层面知识融合框架研究

国外知识融合框架研究起步较早，其中最著名的当是 2001 年完成的 KRAFT 项目[9]，KRAFT 框架的关键在于将多源异构资源转换为统一的知识模型，后来的研究多少都受到这种框架模式的影响。从宏观层面梳理国内外关于知识融合框架的相关研究，本文认为大体可以将其归为两类：一类是方法导向的框架研究，如基于本体和融合规则的知识融合框架；另一类是服务导向的框架研究，如实现智能协助、早期预警的决策支持系统中的融合框架，如表 1 所示：

表 1　知识融合框架的导向

导向	框架
方法导向的框架研究	基于本体和融合规则的知识融合框架[10]； 基于背景知识的多线索融合框架[11]； 加入分类规则的融合系统[12]； 基于上下文的融合框架[13]； 基于关联数据的知识融合框架[14]； 基于概率论的知识融合框架[15]
服务导向的框架研究	实现智能协助、早期预警的决策支持系统中的融合框架[16]； 数字参考咨询知识融合框架[17]； 用于提高云制造效率的知识融合框架[18]； 用于精准农业的知识融合框架[19]； 支持健康评估管理的知识融合框架[20]； 用于边缘领域知识验证的知识融合框架[21]

2.2　微观层面知识融合框架研究

由数据融合和信息融合继承发展而来的知识融合，是个更复杂的过程。从微观层面来看，几乎所有的知识融合框架都体现了知识融合的层级划分。纵向视角上，知识融合框架被划分为不同的层次；横向视角上，知识融合被划分为不同的维度，如表 2 所示：

表 2　知识融合框架的层次与维度

层次与维度	框架
分层融合(纵向)	融合前、融合中、融合后[22]； 基础库、关联级融合、特征级融合、需求级融合[23]； 数据层、模型层、应用层[24]； 知识资源层、知识组织层、知识服务层[25]
多维融合(横向)	知识层融合、方法层融合和思想层融合[26]； 属性融合、实例融合和概念融合[27]； 数据层融合、模型层融合以及参数层融合[12]； 简单融合、扩展融合、实例融合、参数融合、适应性融合、分面融合和历史融合[13]； 内容维度、结构维度、应用维度[28]； 实例融合、域集融合、关系融合、属性融合和概念融合[29]

可以看出，分层融合中层次是纵向视角下的，体现了知识融合过程的不同阶段。虽然知识融合所分的层次名称有所不同，但本质上却有很多相似之处。经过梳理总结，可以进一步将其分为 3 个层次，即数据层、融合层和应用层。数据层融合的主要目的是借助本体、知识元等方法将数据或信息资源进行关联，形成统一的知识表示。融合层融合的主要目的是识别数据层中的概念关系，形成有效的组织体系，并利用推理、融合算法形成新知识。应用层的融合是为了使融合的结果与用户的需求相匹配，并实现有效的知识服务。

如果将分层融合看成是纵向的，那么多维融合无疑是横向的。知识包括实例、属性、概念、关系等，多维融合体现了知识的不同方面，对其进行融合，就构成了不同维度上的知识融合。

2.3　面向大数据的知识融合框架研究

目前面向大数据的知识融合框架研究主要集中于国内。王曰芬等[30]研究设计了大数据环境下知识融合框架的结构和构成要素，包括功能与应用体系、运行与保障体系以及评估与优化体系。唐晓波等[31]提出的大数据环境下知识融合框架模型由数据获取与知识表示、统一知识模式构建、融合处理和衍生知识处理 4 部分构成。张心源等[32]结合已有知识融合框架、系统以及算法等研究基础和大数据环境知识资源、技术方法的变革，设计了大数据环境下以提供更为精准化知识服务为融合最终应用目标的知识融合框架，具体包括融合预处理、融合过程和融合服务 3 个关键部分。

大数据环境具有实时变化的特点，因此实现知识的动态融合也成为了新的研究热点。苏新宁等[33]通过改进知识融合算法，建立与知识库词条对应的动态知识点生成的知识链接数据库，保证动态知识点能够快速生成。L. Tan 等[34]提出了动态功能模型工具和

知识融合方法，以实现在线消费者参与的有效性、高效性。

根据融合的对象不同，知识融合可分为决策级、特征级和数据级 3 个等级[35]。决策级和特征级可称为高级融合，数据级则为低级融合。随着级别的依次降低，融合的数据量和难度依次加大。早期受到计算能力和方法的限制，研究主要集中于高级融合。近年来，随着大数据技术的发展和需求的变化，研究逐渐偏向于低级融合。这种低级融合，在大数据时代反映在对全数据进行分析，而不是小样本分析。周丽琴等[29]在面向网络大数据知识融合框架中，对网络大数据进行了采集、抽取和知识表示。朱娟等[36]在所构建的三层知识模型中融合了大数据环境下的多源信息，实现了商品的个性化推荐。

3 知识融合过程

由于知识本身的复杂性，难以将知识直接进行融合，统一形式化的知识表示是实现知识融合的前提，因此，知识融合的基本思路是首先要对知识进行统一形式化表示。目前知识融合所采取的知识表示形式主要包括知识元和知识网络。据此，本文将知识融合进一步分为知识元融合和知识网络融合，并对这两种形式的知识融合过程进行解析。

3.1 知识元融合

目前情报学领域多数知识融合研究中对知识的统一表示是基于知识元(knowledge element, KE)的。知识元融合的主要思想是首先将各种类型的数据源转换为知识元作为融合的数据基础，然后在本体的支持下，使用算法和规则等融合知识元，以获得可以利用的新知识。基本的融合过程，如图 1 所示：

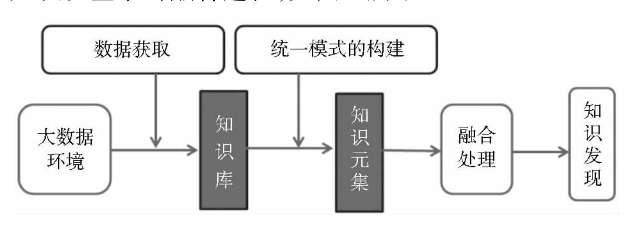


图 1 知识元融合过程示意

国内外许多学者对知识元进行了研究，但目前尚未形成统一的定义，对于其形式化的表示也各有提法。通常，知识元是指能够在语义上完整表达知识的不可分割的最小单元[37]。知识元是“关于知识的知识”，可以描述知识的结构化特征[38]。假定知识元 A，则有：$A=\{(P_1,R_1),(P_2,R_2),\cdots,(P_j,R_j),\cdots,(P_n,R_n)\}$，其中 P 表示知识元属性特征，R 表示属性特征的值。由于知识的复杂性，难以对知识本身进行融合，在本体层面进行操作也同样困难。知识元是对知识的结构化表示，为知识融合提供了合适的数据基础。

在国内知识融合研究中缑锦[39]最先使用了知识元理论，他在研究中对分布式知识库进行了知识元集的转化与处理，构建了数据的基础层。此后，特别是大数据时代到来后，不少学者提出了大数据环境下以知识元为基础的知识融合框架体系。L. Sun 等[40]探索了一种用于 SWOT 情境评估的竞争情报知识融合方法，使基于知识元的隐性智能挖掘成为可能。唐晓波等[41-42]所提出的知识融合框架，是指从大数据集中进行数据获取与表示，形成知识库，然后构建知识元集对知识进行统一的形式化表示，并借助语义熵对知识元进行测度，从而形成有效的知识元集，继而对其进行融合。孙琳等[43]从多个数据源中将竞争情报以知识元的形式进行了抽取和表示，然后对知识元的属性/关系进行融合，实现了新知识的发现。

3.2 知识网络融合

在大数据环境下的多源异构知识中，不同知识实体之间存在着复杂的联系。而知识网络可以结构化地展示实体间的关系，因此，近年来利用知识网络对真实世界的知识进行表示，并对知识网络进行融合，也成为新的研究热点。知识网络(knowledge network, KN)是由知识节点或知识元素与知识关联构成的一种网状结构，知识库是知识网络的主要载体。如图 2 所示，知识网络融合的主要思路是从知识库或网络中获取的知识元素或知识网络与已有知识网络进行融合，形成融合后的新知识网络，并对新知识网络进行知识发现。

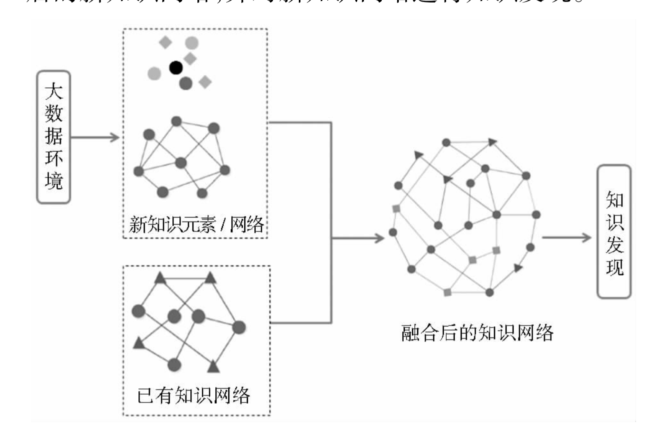


图 2 知识网络融合示意

微软的 Probase[44]和 Google 的 Knowledge Vaule[45]相继提出了从网络大数据中获取知识并进行融合形成

知识网络的模式。林海伦等[46]从实体扩充、关系扩充、分类扩充 3 个方面总结了将新的知识融合到已有知识网络当中的方法。周利琴等[29]从非结构化数据源中抽取相应的知识元素,并将其融合到 Disease Ontology 中的 hypertension 本体中,对该高血压本体进行扩充,以期形成可以解决具体领域问题的知识库。田鹏伟等[47]认为异构信息网络融合方法分为基于元路径提取、多重关系网络及超图/超网络建模等 5 种类型。

4 知识融合方法

知识融合的方法是实现知识融合的关键。目前主要的方法可分为基于语义、基于信息融合算法和基于图模型的知识融合方法 3 种。

4.1 基于语义的知识融合方法

知识与信息的区别之一,就是知识具备推理属性,而语义化是实现知识推理的主要途径,因此基于语义的方法是实现知识融合的主要方法之一。基于语义的方法可以细分为基于语义规则和基于本体论的方法。

4.1.1 基于语义规则的方法 基于语义规则(semantic rules)的方法主要依赖已定义的逻辑运算规则,对融合条件进行约束,进而实现知识融合。J. Gou 等[48]描述了一组比较规则,来明确融合条件,同时通过结果格式和语义规则过滤知识对象,从而实现将知识对象分成不同粒度层次,并且指导融合过程,以规避不合逻辑的结果。Z. Feng 等[49]使用语义相似性来实现语义消歧,使提取的初级词汇链之间实现了知识融合。G. Wang 等[50]提出了一种基于语义元素重构的多源流程创新知识融合算法,并提出了相应的语义冲突解决规则,该算法可以有效地支持所贡献知识的初步自动融合。

4.1.2 基于本体论的方法 本体(ontology)具有良好的概念层次结构,可支持逻辑推理,并且能够通过概念模型实现语义层知识的描述,非常适合知识的表达。很多知识融合的框架是基于本体来构建的,往往和知识元一起同属于框架的数据层。在这样的知识融合框架中,利用知识组织理论,通过本体映射和融合,实现知识融合。目前与知识相关的各分支学科都开始使用本体论,尤其是在基于知识内涵或语义层面的研究。

最早将本体引入知识融合框架的是德国 Ontobroker 工程[51]。N. Xie 等[52]基于农业领域本体来实现多源农业信息源的知识融合问题。J. Liu 等[53]通过分析知识需求,提出了一种动态本体构建方法,以实现更有效的知识融合。A. Smirnov 等[54]在针对实现机器人的自组织所提的知识融合方法中,为了提供语义互操作性,构建了参与场景的机器人 RDF 本体,并且使用了本体匹配技术。H. Fan 等[55]基于本体的知识表示方法,研究了知识融合实现模式,提出了面向知识服务需求的两种维度的知识融合过程模型。

4.2 基于信息融合算法的知识融合方法

知识融合是由信息融合发展而来的,因此信息融合算法在知识融合中得到了继承和发展。在知识融合中使用的信息融合算法主要有贝叶斯网络和 D-S 证据理论。

4.2.1 贝叶斯网络 贝叶斯网络(bayesian network)是一种基于概率推理的图形化网络,可用于发现数据间的潜在关系。贝叶斯网络在解决不确定性和不完整性问题上应用广泛,计算简单直接。E. Santos 等[56]将概率模型表示成贝叶斯知识基础,并提出基于贝叶斯的知识融合算法。K. Coussement 等[57]提出了一个贝叶斯决策支持框架,该框架正式地将主观人类专家意见与更客观的组织信息融合在一起。L. Zhang 等[58]采用贝叶斯方法设计知识类别偏好情境,并根据不同情境的特点,得到知识融合偏好。W. C. Yue 等[59]提出了一种基于模糊贝叶斯网络构造多源知识融合推理模型的方法,该方法可以有效地融合和巩固知识。

4.2.2 D-S 证据理论 D-S 证据理论(dempster-shafer evidence theory)[60]是关于证据和可能性推理的理论,可以消除不确定因素。D-S 证据理论是贝叶斯基础上的近一步发展,支持比贝叶斯更弱的条件。R. Yan 等[61]提出了一种基于 D-S 理论的新型知识融合方法,并将其应用于实际软件故障诊断中;G. Peng 等[62]开发了一种基于 D-S 证据理论的知识融合方法,可用于产品开发;L. Sun 等[63]提出了一种多属性融合的方法,扩展了用于组合型证据的 D-S 理论;L. Q. Sun 等[64]应用基于 D-S 证据理论的知识融合方法研究 H5N1 高致病性禽流感的空间分布。

4.3 图模型方法

除了以上方法外,图模型[65]是近年来较为流行的知识融合方法,图模型可以是概率图、主题图或关系图。通过从其他类型数据中获得先验知识,从而为知识分配一个概率,可看做是图上的链路预测问题。Y. J. Wu 等[66]基于墨西哥湾漏油事件创建了一个大型主题图,并且验证了主题图对于知识融合的有效性;G. Levchuk 等[67]提出了概率图融合算法,以支持多源文本媒体的知识融合和推理;G. Koumoutsos 等[68]提出了

一个可扩展的框架，能够图形化呈现所有可用的结构化和非结构化资源的知识，实现二者的知识融合。

5　知识融合研究趋势

通过对以上研究主题的总结，本文认为当前知识融合研究存在以下特点：①随着大数据的发展和信息技术的进步，本体、语义网、多 Agent、开放关联数据、XML、RDF、Hadoop、MapReduce 等跨学科领域的工具方法，已逐渐被应用到知识融合研究中。多数知识融合框架都可以处理多源异构数据，并且依赖于能够适应各种类型数据的知识抽取和统一的知识表示。②知识元和知识网络作为能够对知识进行结构化表示的方法，具有较强的优越性，已经成为目前知识融合过程中知识表示的主流方法。且知识网络融合在大数据环境下逐渐成为新的研究趋势。③基于语义的方法是知识融合的主要方法，其中本体论方法的研究近年来更加凸显。知识融合还继承了信息融合的算法，并在近年来形成了图模型的方法。

然而当前知识融合研究也存在着以下不足：①虽然学者们提出了不少知识融合框架，但目前尚未形成通用性较强的统一融合框架。尚缺少方法和服务双向视角的、分层多维立体的知识融合框架。②虽然知识元具有很多优越性，但目前知识元的创建过程还需要人工的参与，这影响了知识融合的效率。对多源、异构、大规模数据进行知识网络建模并融合仍存在困难，这是当前制约知识融合向前推进的重要因素。例如在当前的知识网络融合研究中，基于超图超网络方法的融合研究还主要是针对多源异构数据的重新建模，而不是针对大规模异构数据的融合处理。③基于语义规则的知识融合，较大程度地需要人工自定义规则，对于小规模数据量、特定领域还较为有效，但是面对大数据环境，尤其是网络大数据环境，在融合效率和适用性上都有所不足。而且许多知识融合算法依赖于本体互操作，本体的复杂性降低了知识融合的效率，本体的自动化构建也是当前研究的难点。信息融合算法在大数据时代也显得力不从心。贝叶斯网络在面临大规模数据时，产生的网络结构也会较大，因此，在巨大的网络结构中寻找最佳网络会使得存储空间不足，算法效率低下。此算法还要求知识之间的观测独立性、知识先验概率的可预知性，这在大数据环境下都难以满足。D-S 证据理论的融合结果受知识库正确性和完整性的影响较大，在知识融合的决策层或专家知识融合中适用性较高，但不适应于大规模数据的知识融合。而图模型方法基于外部信息源提供的闭合知识集，因此其在大数据环境下的扩展性和适用性不强。总之，以上方法在融合的效率和适用性上还有很大的提升空间。

6　研究展望

自 2012 年大数据概念被正式提出以来，其相关研究以及成果日趋丰富，知识融合的大数据背景已经形成。大数据不仅是一种新概念，更是一种全新的思维方式，不仅为知识融合提供了丰富的数据来源和众多的新技术方法，也带来了多样性、动态性、碎片化以及不确定性等挑战。

大数据环境下，半结构化和非结构化数据规模的飞速增长，使得数据的类型和结构更加复杂。由于自然表达的多样性，存在大量同义和多义的知识表达，造成了知识的语言多样性。大数据环境下的网络、业务系统、数据库中的数据实时更新变化，旧值和新值之间存在冲突。知识对象愈发的碎片化，关系越来越错杂繁复，大量有用的知识分布在各个角落，难以形成系统完整的知识集合。在大数据动态环境中，这种知识的随机性和不确定性也更加明显。因此，多源异构数据的处理更加困难，知识融合遇到了前所未有的挑战。本文认为未来的知识融合研究应关注以下 4 个方面：

6.1　构建分层多维立体的大数据知识融合框架

该框架不仅要能够适应大数据环境，能够处理大规模、多源异构数据，还需满足细粒度、精准化、深层次融合需求。如图 3 所示，大数据环境下的技术和需求不断发生变化，知识融合在分层融合上应更加细化，在多维融合上应更加全面多样，融合级别应更加底层化、全数据化。基于“分层 + 多维 + 全数据”立体架构的知识融合框架，将引领大数据时代的知识融合迈进深度融合阶段。本文认为深度融合就是指对知识进行全方位多角度的融合，挖掘知识之间的深层次关系，实现隐性知识的显性化，并发现新知识，以满足用户的深层次需求。

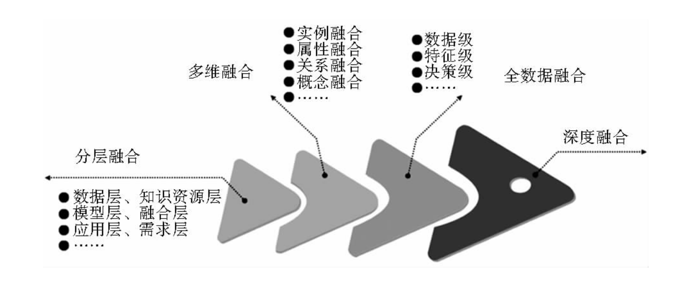


图 3　知识融合框架——以分层、多维和全数据为视角

6.2 提高知识融合的效率

未来应开发更加自动化、结构化的知识表示方法，包括知识元和知识网络的自动化构建，并且减少对本体的依赖。当前如何能够尽可能多地利用本体的现有优势为知识融合服务，同时又能最大程度地克服本体技术中的固有难点，成为一个亟待解决的问题。未来还需开发更为高效精准的知识融合算法。虽然目前的知识融合算法能够适应知识来源的多样性，但还应进一步评估知识本身的真实性和可靠性，并且对知识的结构进行深入、清晰的分析。

6.3 构建实时动态融合机制

在大数据环境，尤其是网络大数据环境下，知识时刻发生变化，并且不同来源知识的更新频率还不一致。这种动态性不仅体现在随时间而变化，还体现在不同的人、不同的外界环境、不同的背景和经验所导致的知识变化。而当前的知识融合方法主要是针对静态知识，对动态知识的融合还处于探索阶段。未来在融合过程中不但要加入知识的时间信息，开发动态融合方法，而且要构建能够实时动态反馈调整的融合机制，以应对大数据环境的动态性挑战。

6.4 开展大数据实证应用研究

在当前的知识融合研究中，知识融合框架、方法研究较多，实证研究较少。在少有的实证研究中，数据规模小，数据的异构复杂程度小，使得现有融合方法在实现大数据环境知识融合的有效性上还有待进一步证明。因此，迫切需要开展真正意义上的大数据知识融合实证研究。

参考文献：

[1] WALTZE E, LINAS J. Intelligence fusion pushed[J]. Aviation week and space technology, 1979(1): 205-211.

[2] TAHANI H, KELLER J M. Information fusion in computer vision using the fuzzy integral [J]. IEEE transactions on systems man & cybernetics, 1990, 20(3): 733-741.

[3] 刘晓娟，李广建，化柏林. 知识融合：概念辨析与界说[J]. 图书情报工作，2016，60(13)：13-19，32.

[4] 祝振媛，李广建. “数据-信息-知识”整体视角下的知识融合初探——数据融合、信息融合、知识融合的关联与比较[J]. 情报理论与实践，2017，40(2)：12-18.

[5] SMIRNOV A, PASHKIN M, CHILOV N, et al. Multi-agent architecture for knowledge fusion from distributed sources[C]//International workshop of central and eastern Europe on multi-agent systems. Heidelberg: Springer, 2001: 293-302.

[6] MARTENS D, DE BACKER M, HAESEN R, et al. Ant-based approach to the knowledge fusion problem[C]//International workshop on ant colony optimization and swarm intelligence. Heidelberg: Springer, 2006: 84-95.

[7] HU X, HU J, SEKHARI A, et al. A fuzzy knowledge fusion framework for terms conflict resolution in concurrent engineering [J]. Concurrent engineering, 2011, 19(1): 71-84.

[8] 邱均平，余厚强. 知识科学视角下国际知识融合研究进展与趋势[J]. 图书情报工作，2015，59(8)：126-132，148.

[9] REECE A, HUI K, GRAY A, et al. Designing for scalability in a knowledge fusion problem[C]//International workshop on ant colony optimization and swarm intelligence. Heidelberg: Springer, 2006: 84-95.

[10] HU X, HU J, SEKHARI A, et al. A fuzzy knowledge fusion framework for terms conflict resolution in concurrent engineering [J]. Concurrent engineering, 2011, 19(1): 71-84.

[11] FONTANI M, ARGONES-RUA E, TRONCOSO C, et al. The watchful forensic analyst: multi-clue information fusion with background knowledge[C]// IEEE International Workshop on Information Forensics and Security. Guangzhou: IEEE, 2013: 120-125.

[12] FISCH D, KALKOWSKI E, SICK B. Knowledge fusion for probabilistic generative classifiers with data mining applications [J]. IEEE transactions on knowledge and data engineering, 2014, 26 (3): 652-666.

[13] SMIRNOV A, LEVASHOVA T, SHILOV N. Patterns for context-based knowledge fusion in decision support systems[J]. Information fusion, 2015, 21: 114-129.

[14] 高劲松，梁艳琪. 关联数据环境下知识融合模型研究[J]. 情报科学，2016，34(2)：50-54.

[15] TAI C H, CHANG C T, CHANG Y S. Hybrid knowledge fusion and inference on cloud environment[J]. Future generation computer systems, 2018, 87: 568-579.

[16] KRIEGEL E U, PFENNIGSCHMIDT S, ZIEGLER H G. Practical aspects of the use of a knowledge fusion toolkit in safety applications[C]// IEEE eleventh international symposium on autonomous decentralized systems. Mexico: IEEE, 2013: 1-4.

[17] 沈旺，李亚峰，侯昊辰. 数字参考咨询知识融合框架研究[J]. 图书情报工作，2013，57(19)：139-143.

[18] LIU J, XU W, ZHAN H. Multi-source and heterogeneous knowledge organization and representation for knowledge fusion in cloud manufacturing[C]//Proceedings of international conference on soft computing techniques and engineering application. New Delhi: Springer, 2014: 55-61.

[19] KAWTRAKUL A, KHUNTHONG V, SUKTARACHAN M, et al. Development of an information integration and knowledge fusion platform for spatial and time based advisory services: precision farming as a case study[C]//Annual SRII global conference. San Jose: IEEE, 2014: 241-248.

[20] YUE K, QIAN W, FU X, et al. Qualitative probabilistic network-based fusion of time-series uncertain knowledge[J]. Soft computing, 2015, 19(7): 1953-1972.

[21] LI F, DONG X L, LANGEN A, et al. Knowledge verification for

long - tail verticals [J]. Proceedings of the VLDB endowment, 2017, 10(11): 1370 - 1381.
[22] WANG F, FAN H, LIU G. Big data knowledge service framework based on knowledge fusion[C]//Proceedings of the international joint conference on knowledge discovery, knowledge engineering and knowledge management. Porto: Lda, 2016: 116 - 123.
[23] 李立睿,邓仲华."互联网 +"背景下科研用户的小数据融合研究[J]. 图书情报工作,2016, 60(6):58 - 63.
[24] 朱娟,唐晓波. 基于三层知识融合模型的个性化商品推荐[J]. 图书馆学研究,2017(5):24 - 30.
[25] 黄新平. 政府网站信息资源多维语义知识融合研究[D]. 长春:吉林大学,2017.
[26] 周芳, 刘玉战, 韩立岩. 基于模糊集理论的知识融合方法研究[J]. 北京理工大学学报(社会科学版), 2013, 15(3): 67 - 73.
[27] 谢能付. 基于农业本体和融合规则的知识融合框架研究[J]. 安徽农业科学,2013,41(1): 395 - 397.
[28] 陈为东, 王萍, 王益成, 等. 政府网站信息资源的多维语义知识融合结构体系及策略研究[J]. 情报理论与实践, 2017, 40(6): 111 - 116.
[29] 周利琴,范昊,潘建鹏. 网络大数据中的知识融合框架研究[J]. 情报杂志,2018,37(1):145 - 150,197.
[30] 王曰芬, 岑咏华. 大数据时代知识融合体系架构设计研究[J]. 数字图书馆论坛, 2016 (10): 16 - 24.
[31] 唐晓波, 朱娟, 杨丰华. 大数据环境下的知识融合框架模型研究[J]. 图书馆学研究, 2016 (1): 32 - 35.
[32] 张心源, 邱均平. 大数据环境下的知识融合框架研究[J]. 图书馆学研究, 2016 (8): 66 - 70.
[33] 苏新宁. 面向知识服务的知识组织理论与方法[M]. 北京:科学出版社, 2014.
[34] TAN L, TANG D, WANG Q, et al. Open design pattern, method, and its self-organization mechanism[J]. Procedia CIRP, 2016, 56: 34 - 39.
[35] 刘渭滨, 邹智元, 邢薇薇. 模式分类中的特征融合方法[J]. 北京邮电大学学报, 2017, 40(4):1 - 8.
[36] 朱娟,唐晓波. 基于三层知识融合模型的个性化商品推荐[J]. 图书馆学研究,2017(5): 24 - 30.
[37] 索传军,盖双双. 知识元的内涵、结构与描述模型研究[J]. 中国图书馆学报, 2018, 44(4): 54 - 72.
[38] 温有奎,焦玉英. 基于知识元的知识发现[M]. 西安:西安电子科技大学出版社,2011.
[39] 缑锦. 知识融合中若干关键技术研究[D]. 杭州:浙江大学, 2005.
[40] SUN L. Knowledge element-based competitive intelligence analytics serving for SWOT situation assessment[C]// The 8th international symposium on computational intelligence and design. Hangzhou: IEEE, 2015, 1: 576 - 579.
[41] 唐晓波,朱娟. 大数据环境下知识融合的关键问题研究综述[J]. 图书馆杂志,2017,36(7): 10 - 16.
[42] 唐晓波,魏巍. 知识融合:大数据时代知识服务的增长点[J]. 图书馆学研究,2015(5):9 - 14,8.
[43] 孙琳,王延章. 基于知识元的多源竞争情报融合方法研究[J]. 情报杂志,2017,36(11):65 - 71.
[44] WU W, LI H, WANG H, et al. Probase: a probabilistic taxonomy for text understanding [C]//Proceedings of the 2012 ACM SIGMOD international conference on management of data. Scottsdale: ACM, 2012: 481 - 492.
[45] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: a Web-scale approach to probabilistic knowledge fusion[C]//Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2014: 601 - 610.
[46] 林海伦, 王元卓, 贾岩涛, 等. 面向网络大数据的知识融合方法综述[J]. 计算机学报, 2017, 40(1): 1 - 27.
[47] 田鹏伟, 张娴, 胡正银, 等. 异构信息网络融合方法研究综述[J]. 图书情报工作, 2017, 61(7): 137 - 144.
[48] GOU J, JIANG Y, WU Y, et al. A new knowledge fusion method based on semantic rules[C]// The 8th international conference on Signal processing. Beijing: IEEE, 2006: 1865 - 1868.
[49] FENG Z, LI-JUAN Z, XIAO-MING F, et al. The research of Vietnamese language news clue extraction method based on converged network semantic knowledge[C]//The 27th Chinese conference on control and decision. Qingdao: IEEE, 2015: 3287 - 3289.
[50] WANG G, HU Y, TIAN X, et al. An integrated open approach to capturing systematic knowledge for manufacturing process innovation based on collective intelligence[J]. Applied sciences, 2018, 8(3): 340.
[51] FENSEL D, DECKER S, ERDMANN M, et al. Ontobroker: how to make the WWW intelligent[C] //Proceedings of the 11th Banff knowledge acquisition for knowledge-based systems. Banff: AAAI, 1998:36 - 42.
[52] XIE N, WANG W, YANG X, et al. Rule - based agricultural knowledge fusion in Web information integration[J]. Sensor letters, 2012, 10(1/2): 635 - 638.
[53] LIU J, XU W, JIANG H. Research on dynamic ontology construction method for knowledge fusion in group corporation[J]. Advances in intelligent systems & computing, 2014, 278:289 - 298.
[54] SMIRNOV A, KASHEVNIK A, MIKHAILOV S, et al. Multi-level robots self-organization in smart space: approach and case study [C]//Conference on smart spaces. Cham: Springer, 2015: 68 - 79.
[55] FAN H, WANG F, ZHENG M. Research on knowledge fusion connotation and process model[C]//China conference on knowledge graph and semantic computing. Singapore: Springer, 2016: 184 - 195.
[56] SANTOS JR E, WILKINSON J T, SANTOS E E. Fusing multiple bayesian knowledge sources[J]. International journal of approximate reasoning, 2011, 52(7): 935 - 947.

[57] COUSSEMENT K, BENOIT D F, ANTIOCO M. A bayesian approach for incorporating expert opinions into decision support systems: a case study of online consumer-satisfaction detection[J]. Decision support systems, 2015, 79: 24 – 32.
[58] ZHANG L, QIAN W, ZHANG Z, et al. Research on knowledge demand information acquisition for product design[C]//The 12th international conference on computational intelligence and security. Wuxi: IEEE,2016: 677 – 680.
[59] YUE W, CHEN X, GUI W, et al. A knowledge reasoning fuzzy-bayesian network for root cause analysis of abnormal aluminum electrolysis cell condition[J]. Frontiers of chemical science and engineering, 2017, 11(3): 414 – 428.
[60] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping[J]. Annals of mathematical statistics, 1967, 38(2):325 – 339.
[61] YAN R, LI G, LIU B. Knowledge fusion based on DS theory and its application on expert system for software fault diagnosis[C]// Prognostics and system health management conference. Beijing: IEEE,2015: 1 – 5.
[62] PENG G, MAO H, WANG H, et al. BOM-based design knowledge representation and reasoning for collaborative product development[J]. Journal of systems science and systems engineering, 2016, 25(2): 159 – 176.
[63] SUN L, WANG Y. A multi-attribute fusion approach extending dempster-shafer theory for combinatorial-type evidences[J]. Expert systems with applications, 2018, 96: 218 – 229.
[64] SUN L, WARD M P, LI R, et al. Global spatial risk pattern of highly pathogenic avian influenza H5N1 virus in wild birds: a knowledge-fusion based approach[J]. Preventive veterinary medicine, 2018, 152: 32 – 39.
[65] LAO N, MITCHELL T, COHEN W W. Random walk inference and learning in a large scale knowledge base[C]//Proceedings of the conference on empirical methods in natural language processing. Stroudsburg: Association for Computational Linguistics,2011: 529 – 539.
[66] WU Y, LEHMAN A, DUNAWAY D J. Evaluations of a large topic map as a knowledge organization tool for supporting self-regulated learning[J]. Knowledge organization, 2015, 42(6): 386 – 398.
[67] LEVCHUK G, BLASCH E. Probabilistic graphical models for multi-source fusion from text sources[C]// Symposium on computational intelligence for security and defense applications. Verona: IEEE, 2015: 1 – 10.
[68] KOUMOUTSOS G, FASLI M, LEWIN I, et al. Graph-based information exploration over structured and unstructured data[C]// IEEE international conference on big data. Boston: IEEE, 2017: 1991 – 2000.

作者贡献说明:
朱祥:提出研究思路和论文框架,撰写论文并修改;
张云秋:确定论文选题,提出修改意见,完善和修改研究内容。

Progress and Trend of Knowledge Fusion Research in Recent Years

Zhu Xiang Zhang Yunqiu
Department of Medical Informatics, School of Public Health, Jilin University, Changchun 130021

Abstract: [**Purpose/significance**] Knowledge fusion, as an effective method to deal with multi-source heterogeneous data and generate new knowledge semantically, has become a new research point of information science in the big data environment, but it is still in its infancy. The paper aims to sort out, evaluate the current research on knowledge fusion, and provide reference for future research. [**Method/process**] Firstly, the concept of knowledge fusion was analyzed. Then the framework, process and method of knowledge fusion were combed. Then the research trend of knowledge fusion was summarized. Finally, the research prospect was made. [**Result/conclusion**] Knowledge fusion research presented new research characteristics in the big data environment, but it can't meet the requirements of the big data environment. In the future, we should build a hierarchical and multi-dimensional big data knowledge fusion framework, improve the efficiency of knowledge fusion, build real-time dynamic fusion mechanism, and carry out big data empirical research based on knowledge fusion.

Keywords: knowledge fusion big data review